Cardiotocography (CTG) refers to the recording of fetal heartbeat and the uterine contractions. Acidosis refers to increased acidity in the blood and other body tissue of a human. In this report, we have used a CTG dataset to predict acidosis in neonates. The models used were Logistic Regression (13 variables), KNN (25 variables), and Classification Tree (25 variables). 352 patients were diagnosed for our CTG data set. The data set contained 2126 rows along with 40 variables. Out of the 40 variables, 16 were categorical and 24 numerical. Of the numerical variables, 18 were calculations performed from original values recorded by the cardiotocograph - the machine used to perform the monitoring of the fetal heart rate and uterine contractions. Exhibit A shows the number and type of decelerations versus the number of unique fetuses. The 4 variables are DL (Light Decelerations), DP (Prolonged Decelerations), DR (Repetitive Decelerations), and DS (Severe Decelerations). All these variables are calculations that originate from the fetus' heart rate (FHR). To understand this, please refer to Exhibit S . Each deceleration in FHR is recorded as 1 unit in our CTG dataset. For instance, Exhibit A indicates that there were 15 fetuses with light decelerations and 2 fetuses with severe decelerations. Other variables in the dataset measured sleep patterns and fetal movement.

The dataset contained no missing values, no outliers, and no duplicates. To find out about any missing values we executed an R command which checks each variable for any missing values as seen in Exhibit B. To find about outliers we performed boxplots on every variable as seen in Exhibits D & E which show boxplots of two different variables: Width of the Histogram (Width) and Fetal Heart Rate (LBE). Finally to find out about duplicates we executed a function in excel which checked for any duplicates as shown in Exhibit C.

The response in our dataset is NSP referring to 'Normal' state (1), 'Suspect' state (2), and 'Pathologic' state (3). Each classification intuitively lets the doctor know whether the fetus is healthy, needs to be further diagnosed, or needs to be treated, respectively. In order to perform our predictive models, we had to make some adjustments to our dataset. Particularly, our Binary Logistic Regression demanded that we turn our response into a binary response rather than a multicategorical response as seen in Exhibits F and G. the new binary response would consist of 'Normal' state (1), and 'Abnormal' state (2), meaning 'healthy' and 'not healthy.' To satiate our curiosity, we performed all of our models on the binary response modified dataset as well. It was specially interesting to compare the outcomes of both multicategorical and binary responses in our models, and to analyze the tradeoffs and problems that arose from using the multicategorical version of the model as compared to its binary counterpart.

Other modifications on our dataset consisted of eliminating variables to improve the performance of our models. All redundant variables such as the file name (FileName) and the date of the examination (Date) were eliminated. Variables that contained a very high percentage of 0s (over 90 percent) were also nulled as most of them did not mathematically contribute to the models. Exhibit H is a histogram of the variable DS (Severe Decelerations) which we eliminated do to its extremely high percentage of 0s.

More insignificant variables were deleted such as DR (Repetitive Decelerations) which consisted of 100 percent 0s, as shown in Exhibit I. There were exceptions nonetheless, as we excluded a few variables that had a very high percentage of 0s and yet worsened the model's performance. An example of these kinds of variables are DP (Prolonged Decelerations) and E (Sleep Pattern) shown in Exhibit J and K. We do not know why these variables contributed to

our models and we believe the reason for this significance is directly related to the medical field itself.

**Logistic Model:**

Working on this project was one of the most challenging things we have done throughout the semester.We were asked to get data, import it to R-Studio, run the Macros and interpret the results. The description of the data was not technical and was always provided to us. However, with our project, the Cardiotocography (CTG) dataset that we analyzed had many technical terms and definitions (Bache). In addition, measurements were unclear since the provided Excel sheet didn't provide the units that were used to measure different variables. Only after reading many scholarly articles and other Biomedical Engineering research, we started to understand the meaning of variables and even how to interpret a CTG graph at some extent. There have been many data analysis' performed on this specific dataset and it is even included in different textbooks (Marques). Outside sources helped with understanding the data and what the variables mean.

When we first tackled the Logistic model, we ran it with two different sets of variables. However, after meeting with the instructor, we realized that it needed three responses, which we had not done in class. We tried to look up the Multiple Logistics, but we were afraid we would not get it done on time.

After, we decided to combine two of the responses into one group. The response variable for this project is named (NSP) which stands for Normal=0, Suspect=1, Pathologic=2.
We put Suspect and Pathologic into one group under the categorical outcome 1, keeping Normal = 0. The professor highly recommended we do this so that we could run our model successfully.

Next step was to add a few variables each at a time to the model. We eliminated the insignificant variables, such as the name of the file and the date. The first Logistic model that we ran used five variables that were mostly measurements of the heartbeat (LBE), the Fetal Movement (FM), and Uterine Contractions (UC) along with other variables related to regular and irregular accelerations and decelerations of the heartbeat of the fetus.

Having a better understanding of the variables and throwing few variables into the model one at a time made it easy for our group to come up with different models. The first model had a specificity of 0.98, and the same for the three models that followed. However, the sensitivity was 0.92 and improved up to 0.96 for the final model. A quick look at the Confusion Matrix for the final model shows the final specificity is 0.98 and final sensitivity is 0.96 (EXHIBITS L and M) To understand a few of the variables within the data and how to interpret them, we will explain the Fetal Movement, since it is an somehow easy to explain and can be measured without using fancy devices.

The parameter estimate corresponding to the Fetal Movement (FM) is statistically significant and equals 0.014. That means, that **when holding all the predictors other than FM constant**, per 1 unit increase in FM, the odds that the fetus's heartbeat is normal changes by a factor of $e^{0.014} = 1.014$. ( ROC Plot EXHIBIT N)

The increase in the fetal movement is very important because it must be monitored at all times. The American Congress of Obstetricians and Gynecologists (ACOG) recommends that you time how long it takes you to feel 10 kicks, flutter, swishes, or rolls. Ideally, you want to feel at least 10 movements within 2 hours. International standards are the same as defined by the International Association of Obstetricians and Gynecologists. Therefore, this Logistic model can

be applied to any set of data. We believe that this analysis will also be useful to new companies that are trying to build a new device for undeveloped countries, and this analysis can give them the option of adjusting the parameters according to their needs. In the United States, there are almost over 40,000 newborns that die from heart complications. The device that has been used to record this data is fairly expensive, so knowing what kind of data to record and how to run an analysis can save a lot of lives and money.

**Classification Tree**

We created a classification tree using the provided CART Macro. The purpose of this method is to create a visually appealing interpretation of the data so that a medical professional can understand it at a glance. The tree will help classify an observation into one of the pre-defined classes. Since we used both the raw data, which has 3 outcomes (normal, suspect, pathologic), and our own modified data which has 2 outcomes (normal, abnormal), we have 2 different trees/matrices.

In Exhibit O, we have the confusion matrix and the classification tree for our data with 3 outcomes. Starting with the confusion matrix, it is nice to take a look at the classification rates, or the percent of observations that were correctly classified by the model. For Normal fetal acid, the classification rate was 100% , which is great. This means that those patients that aren't at risk are never in doubt and don't need any further testing. For Suspect we got a 89.8% classification rate. This means that almost 90% of patients that could potentially have an issue were classified correctly. For Pathologic we got an 88.6% classification rate. Those patients with a potentially serious issue that would require further treatment were classified at almost 89% accuracy.

To interpret our tree, a deep understanding of some medical terms and methodologies used are required. Starting with SUSP, it should be made clear that it is not the same as our outcome "suspect" from NSP. The SUSP on our tree simply indicates whether the patient moved during sleep or not. A 0 refers to no movement, and a 1 refers to movement during sleep. The colon in the statement SUSP:0 lets the viewer know that if this criteria is met, he/she should move left to the next part of the tree. The variable E is another categorical value related to sleep pattern (0 for regular, and 1 for disturbed). Mean and Max relate to the mean and maximum value of a specific observation on a graph. FS stands for flat sinusoidal which put simply means how flat the graph is. DP is the number of prolonged decelerations per second, so in the example provided, that would be 1. Following these steps would tell you whether the patient is classified as normal, suspect, or pathologic.

Taking a look at the tree and confusion matrix in Exhibit P, we see the results of our modified data. The classification rate for normal went down to 98.8% which is minimal, but obviously not an improvement from our previous model. the classification rate for abnormal (the grouping of suspect and pathological) is 97.9%. This is a very good number. Many of the variables in this tree also appeared in the previous tree. The new ones are LBE and AD. LBE is basically the baseline average heart rate in a 10 minute period and AD (accelerative/decelerative) looks at the pattern of the graph, whether it is increasing or decreasing over time.

The results for this tree are pretty good in general. However, these are people's lives we are talking about, not something like sales. This brought up interesting discussions regarding the ethical decisions that businesses have to consider, even if they are profit motivated.

**KNN**

The K-Nearest-Neighbors predictive model was used to classify the true class of fetal acidosis in two ways. First, we classified it how the cardiologists did, as Normal (1), Suspect (2), or Pathologic (3). Then, we classified the model into two classifications, Normal (1) or Abnormal (2). The two models were used to compare the different misclassification rates to see which model would be best for cardiologists to use.

The difference between the variables used for CART and KNN is that for CART, any variable that only had values of zero and one were converted to categorical variables so that skewness in the model could be prevented. These variables were Calm Sleep (A), Abnormal or Normal sleeping pattern (E), Accelerative/Decelerative Pattern (AD), Decelerative Pattern (DE), Flat Sinusoidal Pattern (FS), and Suspect Pattern (SUSP). All of these variables had to do with the sleeping pattern of the mother and were observed as a 0 (not present), or 1 (present). Since KNN cannot use categorical variables, they were kept as integers for the KNN Macro. Although many of the values for the KNN models had observations of zero or one, the standardization of the variables made it so the model would not be skewed by these values.

Both models took less than a minute to compute the KNN code in RStudio. K=1 was the best value for K for both the model with three categories and the model with two categories. This means that only a minimal amount of information was used since only the closest distance, or 'nearest neighbor', is used to determine the predicted classification. The closer the distance of the observation used to predict the new classification, the more likely the classification will be correct since the 'nearest neighbor' will have such similar features of the new classification. The first model, with 3 categories (Normal, Suspect, and Pathologic), had a misclassification rate of 0.7%. The misclassification matrix is shown in Exhibit Q. The classification rate for Normal

Fetal Acid is 99.7%, Suspect Fetal Acid is 98.3% and Pathologic Fetal Acid is 97.14%. Although this is a fantastic model, with an overall classification of 99.3%, it would be an even better model if the Suspect and Pathologic classification rates were higher than the Normal classification rate. This is because in medicine, it is more important for a doctor to know if the fetal acid is Pathologic, rather than if it is Normal.

Due to the need for a better classification rate when dealing with Suspect and Pathologic fetal acidosis, it was decided to combine the two categories together, creating new categories named Normal (1) and Abnormal (2), to see if it would improve the KNN model. The overall classification rate of this new model with two categories is 99.53%. While this only seems like a slightly better model compared to the previous model, it is better in predicting abnormal fetal acidosis which could help doctors save more lives. The misclassification matrix is shown in Exhibit R. The classification rate for Normal Fetal Acid is the same as before, 99.7%, but the classification rate for Abnormal Fetal Acid is 98.9% which is better than both the Suspect and Pathologic percentages in the first model.

It depends on the doctor as to which model would be best for them. The first model, while not as accurate, would give doctors a more precise representation of the severity of the fetal acidosis. It would most likely be cheaper for the hospital as well, since one would find out the stage of the fetal acidosis from the very beginning and not as many tests would be needed. The second model would be best for any doctor wanting the highest classification rate. This model would save more lives due to its ability to better classify if the fetal acidosis is abnormal.

**Conclusion**

Exhibit T displays the different classification rates achieved by the different models performed throughout the project. All models had a high classification rate but as noted above, in the healthcare field we deal with real lives and we need to be very careful about which model to select The KNN model was the best model of all for both the multicategorical response dataset and the binary response dataset, given that it performed better than the other models in the Suspect and Pathologic states in the case of the multicategorical response dataset, or the Abnormal state in the case of the binary response dataset. This however, does not mean that the other models were not useful. The logistic model provided rich insight into our data and achieved a high classification rate using almost half the variables than the other two models. The Classification Tree model gave us an idea of the most important variables in our dataset thanks to its 10-fold cross-validation, and it provided us with a relatively intuitive picture that can be used by doctors to make quick predictions about acidosis in neonates without using a machine to perform an algorithm.

When looking at the KNN model, The binary model's overall classification rate was slightly better than the multicategorical model's classification rate. Whether the hospital/clinic decides to use the binary model or the multicategorical model is up to them and depends on whether they have the resources and money to diagnose and examine fetuses that would go directly to treatment if classified by the multicategorical model. Above all it is very interesting how the healthcare industry needs to consider both business and ethical decisions of this kind and balance between one or another, even if that means losing lives.

# Exhibits

EXHIBIT A

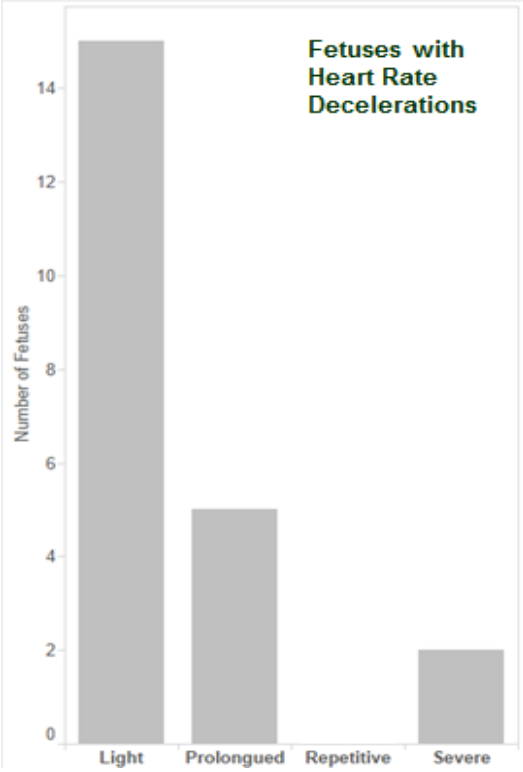
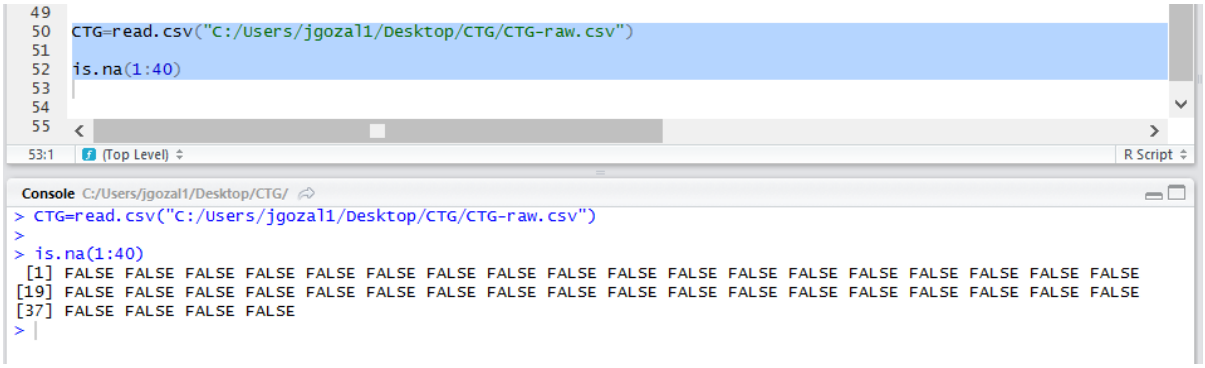
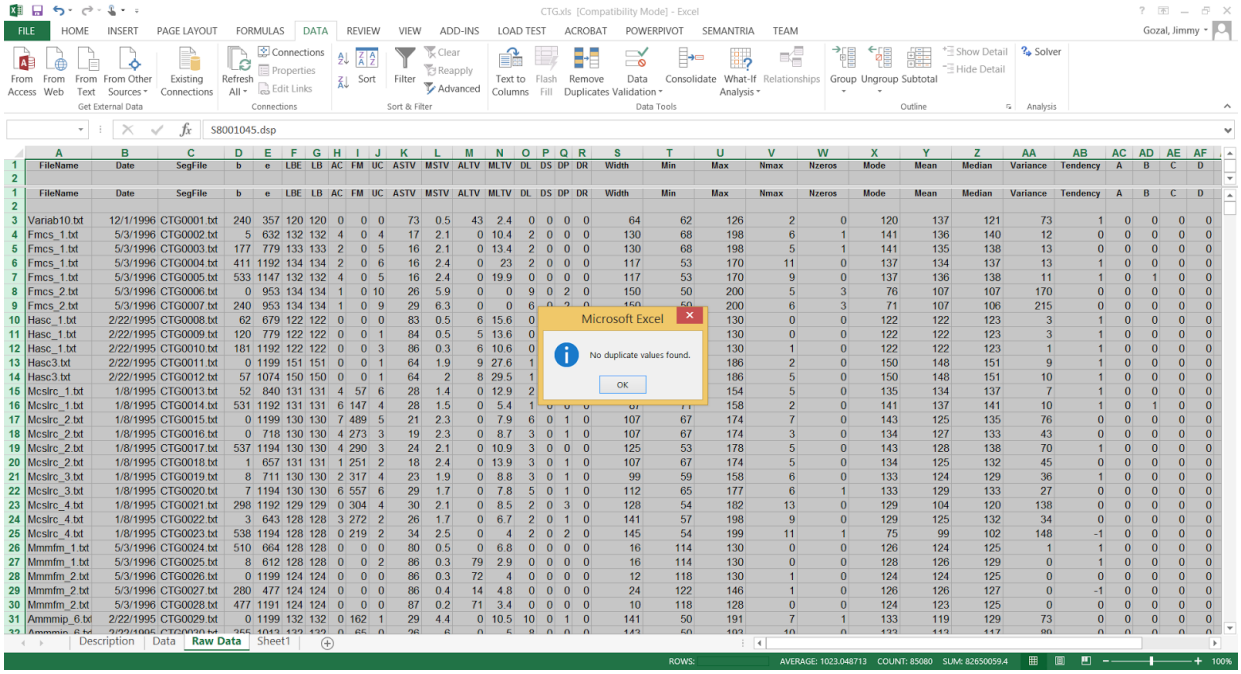



EXHIBIT B



# EXHIBIT C




| FileName | Date | SegFile | b | e | LBE | LB | AC | FM | UC | ASTV | MSTV | ALTV | MLTV | DL | DS | DP | DR | Width | Min | Max | Nmax | Nzeros | Mode | Mean | Median | Variance | Tendency | A | B | C | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variab10.txt | 12/1/1996 | CTG0001.txt | 240 | 357 | 120 | 120 | 0 | 0 | 0 | 73 | 0.5 | 43 | 2.4 | 0 | 0 | 0 | 0 | 64 | 62 | 126 | 2 | 0 | 120 | 137 | 121 | 73 | 1 | 0 | 0 | 0 | 0 |
| Fmcs_1.txt | 5/3/1996 | CTG0002.txt | 5 | 632 | 132 | 132 | 4 | 0 | 4 | 17 | 2.1 | 0 | 10.4 | 2 | 0 | 0 | 0 | 130 | 68 | 198 | 6 | 1 | 141 | 136 | 140 | 12 | 0 | 0 | 0 | 0 | 0 |
| Fmcs_1.txt | 5/3/1996 | CTG0003.txt | 177 | 779 | 133 | 133 | 2 | 0 | 5 | 16 | 2.1 | 0 | 13.4 | 2 | 0 | 0 | 0 | 130 | 68 | 198 | 5 | 1 | 141 | 135 | 138 | 13 | 0 | 0 | 0 | 0 | 0 |
| Fmcs_1.txt | 5/3/1996 | CTG0004.txt | 411 | 1192 | 134 | 134 | 2 | 0 | 6 | 16 | 2.4 | 0 | 23 | 2 | 0 | 0 | 0 | 117 | 53 | 170 | 11 | 0 | 137 | 134 | 137 | 13 | 1 | 0 | 0 | 0 | 0 |
| Fmcs_1.txt | 5/3/1996 | CTG0005.txt | 533 | 1147 | 132 | 132 | 4 | 0 | 5 | 16 | 2.4 | 0 | 19.9 | 0 | 0 | 0 | 0 | 117 | 53 | 170 | 9 | 0 | 137 | 136 | 138 | 11 | 1 | 0 | 1 | 0 | 0 |
| Fmcs_2.txt | 5/3/1996 | CTG0006.txt | 0 | 953 | 134 | 134 | 1 | 0 | 10 | 26 | 5.9 | 0 | 0 | 9 | 0 | 2 | 0 | 150 | 50 | 200 | 5 | 3 | 76 | 107 | 107 | 170 | 0 | 0 | 0 | 0 | 0 | 0 |
| Fmcs_2.txt | 5/3/1996 | CTG0007.txt | 240 | 953 | 134 | 134 | 1 | 0 | 9 | 29 | 6.3 | 0 | 0 | 6 | 0 | 2 | 0 | 150 | 50 | 200 | 6 | 3 | 71 | 107 | 106 | 215 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hasc_1.txt | 2/22/1995 | CTG0008.txt | 62 | 679 | 122 | 122 | 0 | 0 | 0 | 83 | 0.5 | 6 | 15.6 | 0 | | | | | | 130 | 0 | 0 | 122 | 122 | 123 | 3 | 1 | 0 | 0 | 0 | 0 |
| Hasc_1.txt | 2/22/1995 | CTG0009.txt | 120 | 779 | 122 | 122 | 0 | 0 | 1 | 84 | 0.5 | 5 | 13.6 | 0 | | | | | | 130 | 0 | 0 | 122 | 122 | 123 | 3 | 1 | 0 | 0 | 0 | 0 |
| Hasc_1.txt | 2/22/1995 | CTG0010.txt | 181 | 1192 | 122 | 122 | 0 | 0 | 3 | 86 | 0.3 | 6 | 10.6 | 0 | | | | | | 130 | 1 | 0 | 122 | 122 | 123 | 1 | 1 | 0 | 0 | 0 | 0 |
| Hasc3.txt | 2/22/1995 | CTG0011.txt | 0 | 1199 | 151 | 151 | 0 | 0 | 1 | 64 | 1.9 | 9 | 27.6 | 1 | | | | | | 186 | 2 | 0 | 150 | 148 | 151 | 9 | 1 | 0 | 0 | 0 | 0 |
| Hasc3.txt | 2/22/1995 | CTG0012.txt | 57 | 1074 | 150 | 150 | 0 | 0 | 1 | 64 | 2 | 8 | 29.5 | 1 | | | | | | 186 | 5 | 0 | 150 | 148 | 151 | 10 | 1 | 0 | 0 | 0 | 0 |
| Mcslrc_1.txt | 1/8/1995 | CTG0013.txt | 52 | 840 | 131 | 131 | 4 | 57 | 6 | 28 | 1.4 | 0 | 12.9 | 2 | | | | | | 154 | 5 | 0 | 135 | 134 | 137 | 7 | 1 | 0 | 0 | 0 | 0 |
| Mcslrc_1.txt | 1/8/1995 | CTG0014.txt | 531 | 1192 | 131 | 131 | 6 | 147 | 4 | 28 | 1.5 | 0 | 5.4 | 1 | | | | | | 158 | 2 | 0 | 141 | 137 | 141 | 10 | 1 | 0 | 1 | 0 | 0 |
| Mcslrc_2.txt | 1/8/1995 | CTG0015.txt | 0 | 1199 | 130 | 130 | 7 | 489 | 5 | 21 | 2.3 | 0 | 7.9 | 6 | 0 | 1 | 0 | 107 | 67 | 174 | 7 | 0 | 143 | 125 | 135 | 76 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_2.txt | 1/8/1995 | CTG0016.txt | 0 | 718 | 130 | 130 | 4 | 273 | 3 | 19 | 2.3 | 0 | 8.7 | 3 | 0 | 1 | 0 | 107 | 67 | 174 | 3 | 0 | 134 | 127 | 133 | 43 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_2.txt | 1/8/1995 | CTG0017.txt | 537 | 1194 | 130 | 130 | 4 | 290 | 3 | 24 | 2.1 | 0 | 10.9 | 3 | 0 | 0 | 0 | 125 | 53 | 178 | 5 | 0 | 143 | 128 | 138 | 70 | 1 | 0 | 0 | 0 | 0 |
| Mcslrc_2.txt | 1/8/1995 | CTG0018.txt | 1 | 657 | 131 | 131 | 1 | 251 | 2 | 18 | 2.4 | 0 | 13.9 | 3 | 0 | 1 | 0 | 107 | 67 | 174 | 5 | 0 | 134 | 125 | 132 | 45 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_3.txt | 1/8/1995 | CTG0019.txt | 8 | 711 | 130 | 130 | 2 | 317 | 4 | 23 | 1.9 | 0 | 8.8 | 3 | 0 | 1 | 0 | 99 | 59 | 158 | 6 | 0 | 133 | 124 | 129 | 36 | 1 | 0 | 0 | 0 | 0 |
| Mcslrc_3.txt | 1/8/1995 | CTG0020.txt | 7 | 1194 | 130 | 130 | 6 | 557 | 6 | 29 | 1.7 | 0 | 7.8 | 5 | 0 | 1 | 0 | 112 | 65 | 177 | 6 | 1 | 133 | 129 | 133 | 27 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_4.txt | 1/8/1995 | CTG0021.txt | 298 | 1192 | 129 | 129 | 0 | 304 | 4 | 30 | 2.1 | 0 | 8.5 | 2 | 0 | 3 | 0 | 128 | 54 | 182 | 13 | 0 | 129 | 104 | 120 | 138 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_4.txt | 1/8/1995 | CTG0022.txt | 3 | 643 | 128 | 128 | 3 | 272 | 2 | 26 | 1.7 | 0 | 6.7 | 2 | 0 | 1 | 0 | 141 | 57 | 198 | 9 | 0 | 129 | 125 | 132 | 34 | 0 | 0 | 0 | 0 | 0 |
| Mcslrc_4.txt | 1/8/1995 | CTG0023.txt | 538 | 1194 | 128 | 128 | 0 | 219 | 2 | 34 | 2.5 | 0 | 4 | 2 | 0 | 2 | 0 | 145 | 54 | 199 | 11 | 1 | 75 | 99 | 102 | 148 | -1 | 0 | 0 | 0 | 0 |
| Mmmfm_1.txt | 5/3/1996 | CTG0024.txt | 510 | 664 | 128 | 128 | 0 | 0 | 0 | 80 | 0.5 | 0 | 6.8 | 0 | 0 | 0 | 0 | 16 | 114 | 130 | 0 | 0 | 126 | 124 | 125 | 1 | 1 | 0 | 0 | 0 | 0 |
| Mmmfm_1.txt | 5/3/1996 | CTG0025.txt | 8 | 612 | 128 | 128 | 0 | 0 | 2 | 86 | 0.3 | 79 | 2.9 | 0 | 0 | 0 | 0 | 16 | 114 | 130 | 0 | 0 | 128 | 126 | 129 | 0 | 1 | 0 | 0 | 0 | 0 |
| Mmmfm_2.txt | 5/3/1996 | CTG0026.txt | 0 | 1199 | 124 | 124 | 0 | 0 | 0 | 86 | 0.3 | 72 | 4 | 0 | 0 | 0 | 0 | 12 | 118 | 130 | 1 | 0 | 124 | 124 | 125 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mmmfm_2.txt | 5/3/1996 | CTG0027.txt | 280 | 477 | 124 | 124 | 0 | 0 | 0 | 86 | 0.4 | 14 | 4.8 | 0 | 0 | 0 | 0 | 24 | 122 | 146 | 1 | 0 | 126 | 126 | 127 | 0 | -1 | 0 | 0 | 0 | 0 |
| Mmmfm_2.txt | 5/3/1996 | CTG0028.txt | 477 | 1191 | 124 | 124 | 0 | 0 | 0 | 87 | 0.2 | 71 | 3.4 | 0 | 0 | 0 | 0 | 10 | 118 | 128 | 0 | 0 | 124 | 123 | 125 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ammmip_6.txt | 2/22/1995 | CTG0029.txt | 0 | 1199 | 132 | 132 | 0 | 162 | 1 | 29 | 4.4 | 0 | 10.5 | 10 | 0 | 1 | 0 | 141 | 50 | 191 | 7 | 1 | 133 | 119 | 129 | 73 | 0 | 0 | 0 | 0 | 0 |

# EXHIBIT D

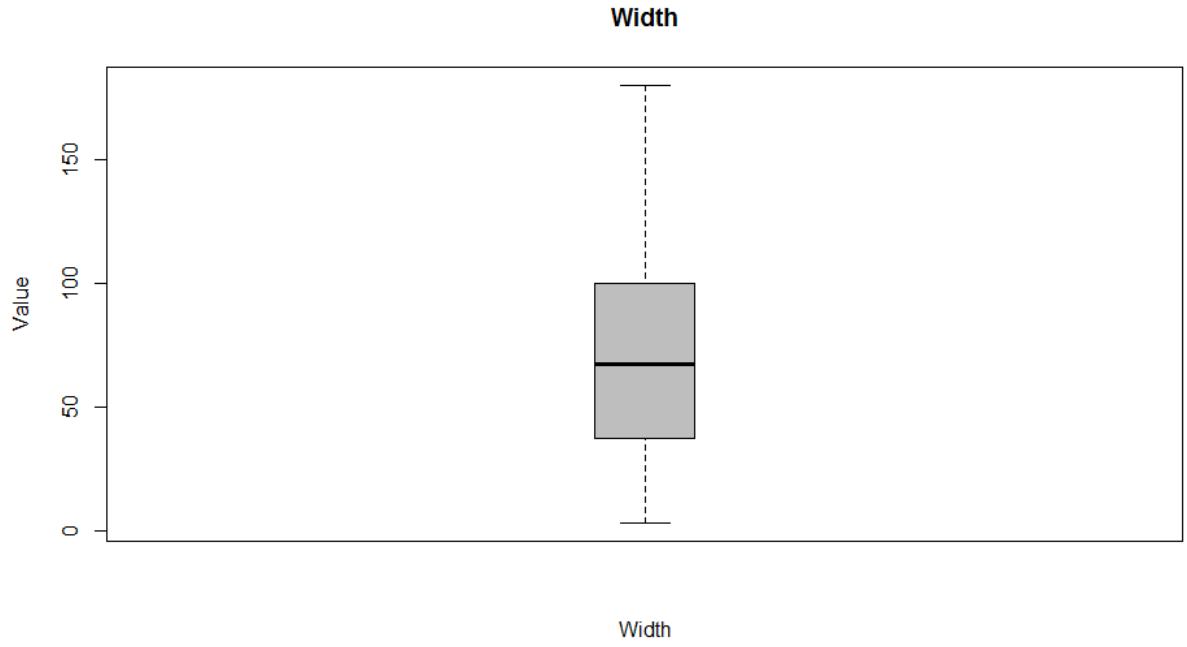


EXHIBIT E

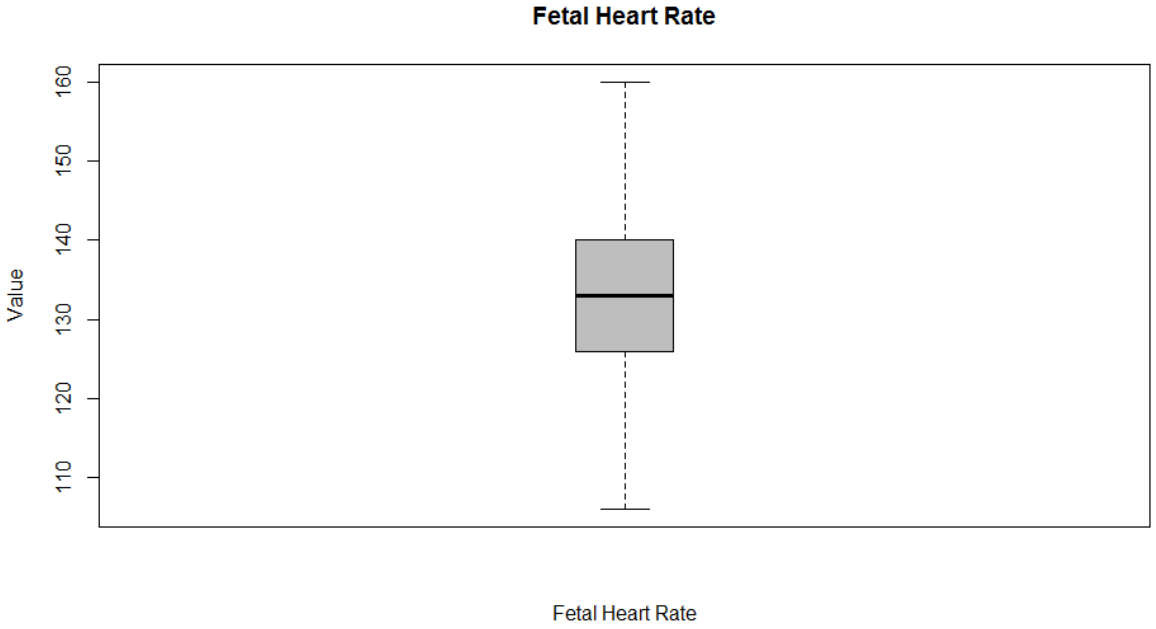



EXHIBIT F

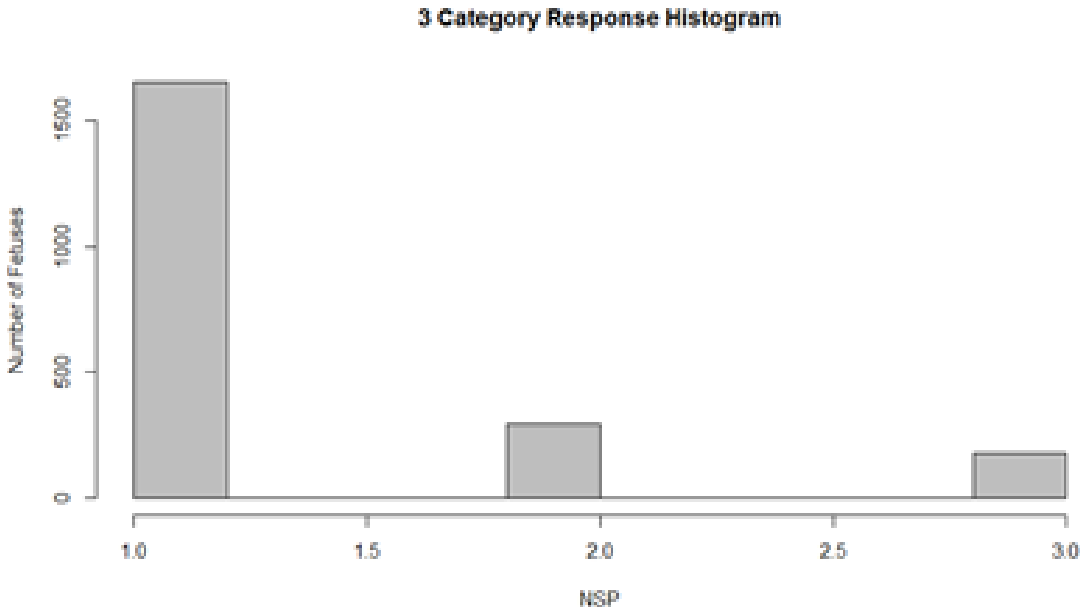


EXHIBIT G

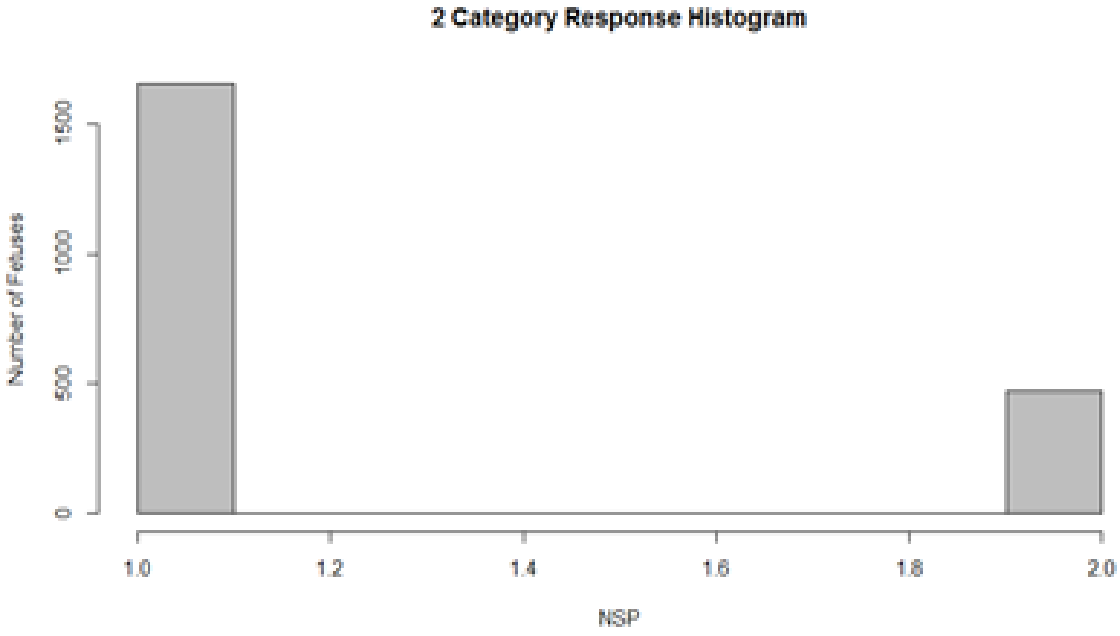



EXHIBIT H

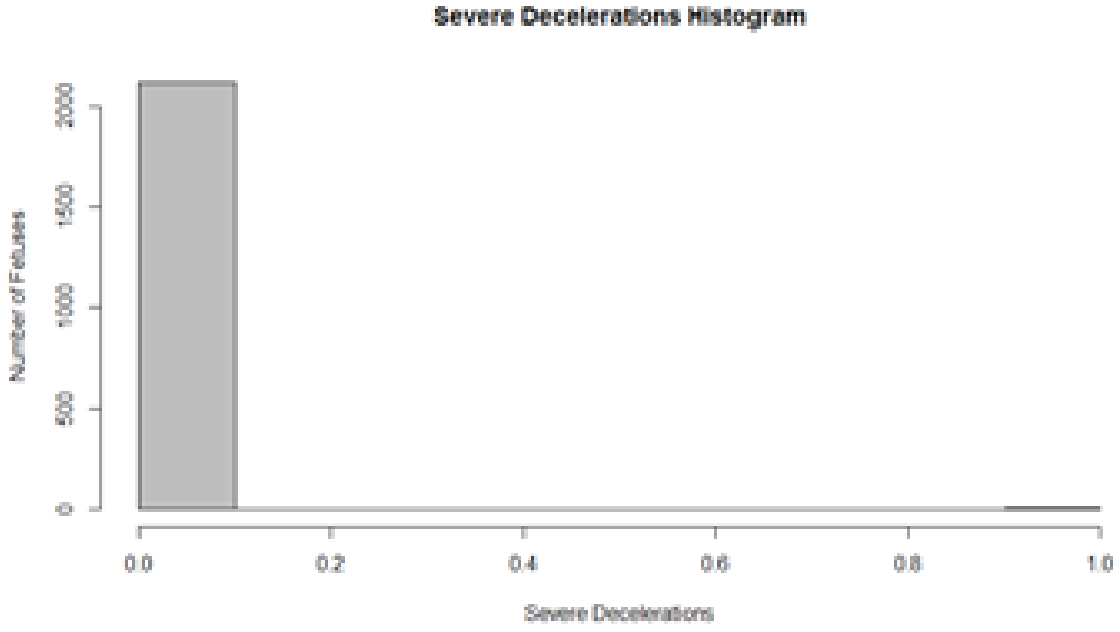


EXHIBIT I

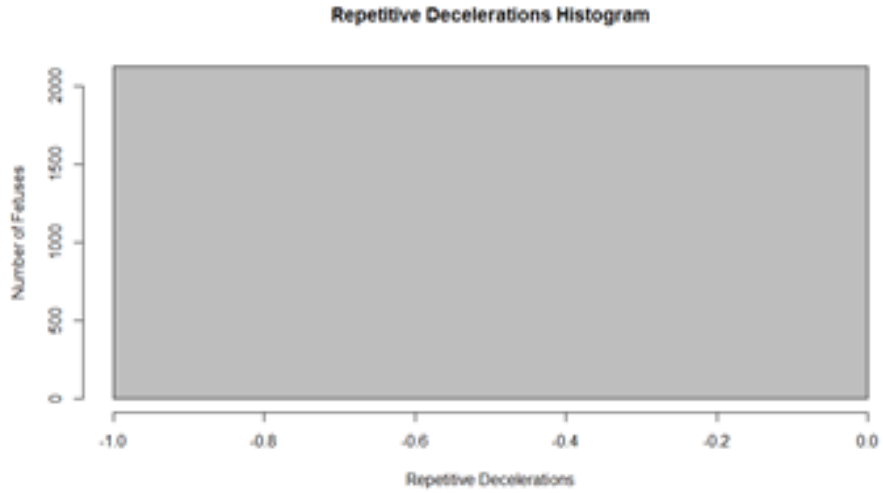



EXHIBIT J

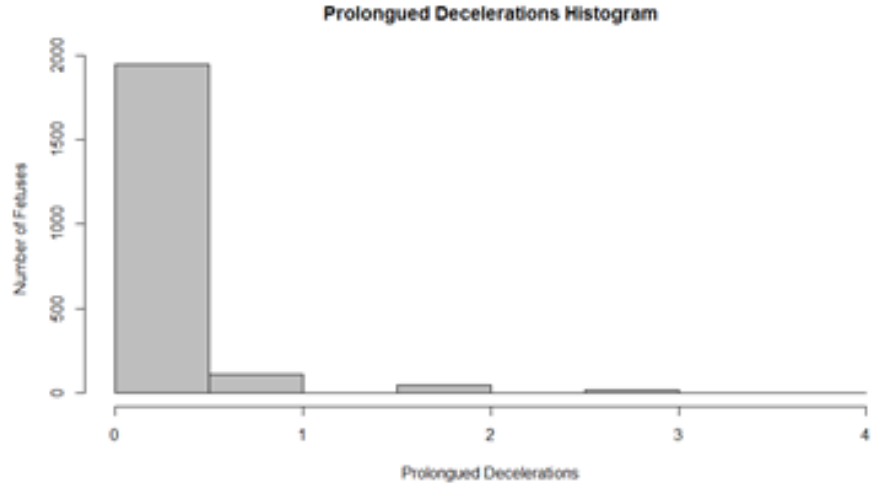



EXHIBIT K

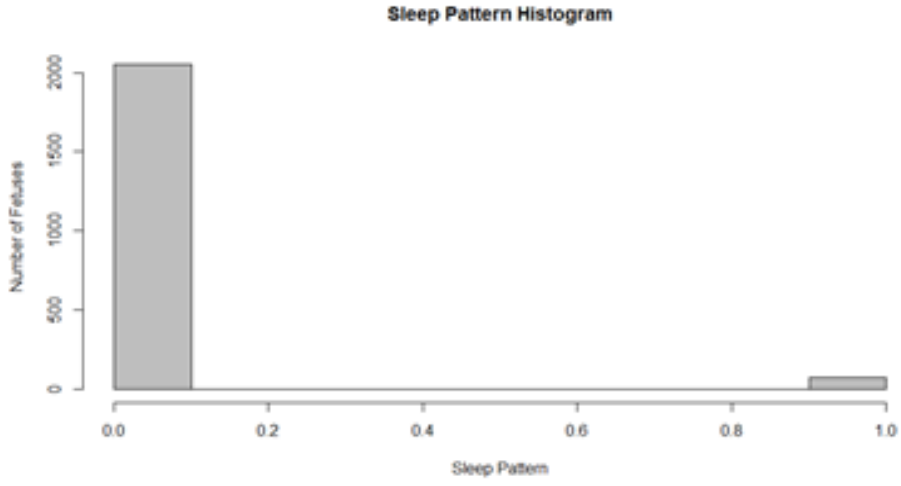


EXHIBIT L

Total Observations in Table: 425

| | Predicted Class | | |
|---|---|---|---|
| True Class | 0 | 1 | Row Total |
| 1 | 326 | 5 | 331 |
| 2 | 4 | 90 | 94 |
| Column Total | 330 | 95 | 425 |

EXHIBIT M

Coefficients:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | |
|---|---|---|---|---|---|
| (Intercept) | -13.393073 | 2.772066 | -4.831 | 1.36e-06 | *** |
| LBE | 0.043253 | 0.019886 | 2.175 | 0.029630 | * |
| AC | -1.291677 | 0.155921 | -8.284 | < 2e-16 | *** |
| FM | 0.014771 | 0.002783 | 5.308 | 1.11e-07 | *** |
| UC | -0.132398 | 0.066133 | -2.002 | 0.045283 | * |
| ASTV | 0.134855 | 0.015407 | 8.753 | < 2e-16 | *** |
| ALTV | 0.066312 | 0.012485 | 5.311 | 1.09e-07 | *** |
| DL | -0.434587 | 0.083689 | -5.193 | 2.07e-07 | *** |
| DP | 2.916606 | 0.414815 | 7.031 | 2.05e-12 | *** |
| Variance | 0.042135 | 0.007191 | 5.860 | 4.64e-09 | *** |
| Tendency | -0.977280 | 0.296324 | -3.298 | 0.000974 | *** |
| A | -8.020173 | 1.068898 | -7.503 | 6.23e-14 | *** |
| E | 5.502647 | 0.888083 | 6.196 | 5.79e-10 | *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

EXHIBIT N

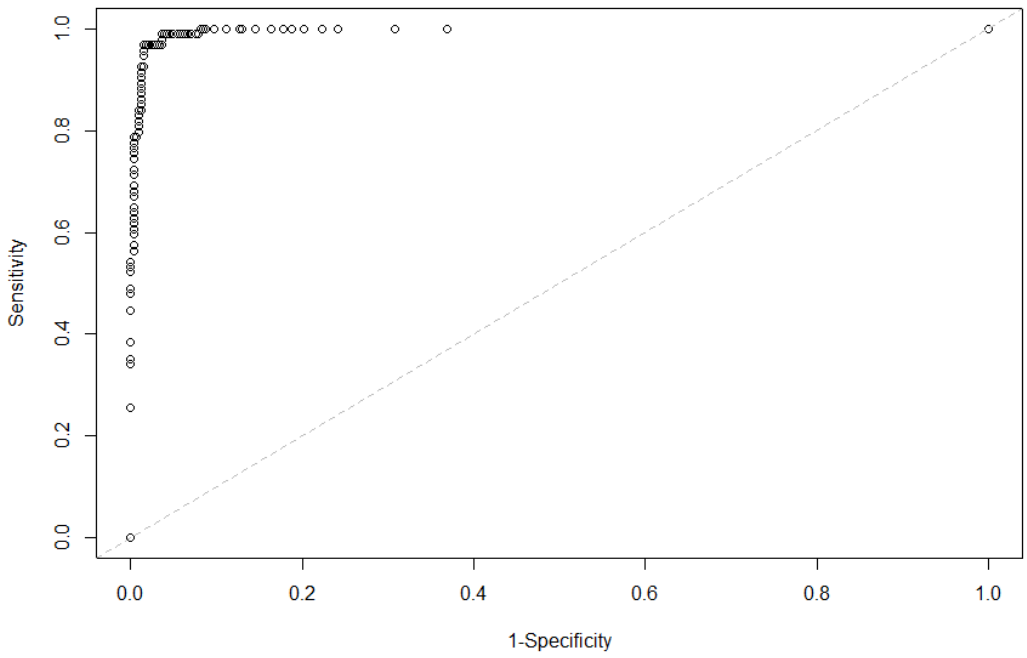


EXHIBIT O

| final_table$myresponse | final_table$predicted 1 | 2 | 3 | Row Total |
|---|---|---|---|---|
| 1 | 331 | 0 | 0 | 331 |
| | 1.000 | 0.000 | 0.000 | 0.779 |
| | 0.971 | 0.000 | 0.000 | |
| 2 | 6 | 53 | 0 | 59 |
| | 0.102 | 0.898 | 0.000 | 0.139 |
| | 0.018 | 1.000 | 0.000 | |
| 3 | 4 | 0 | 31 | 35 |
| | 0.114 | 0.000 | 0.886 | 0.082 |
| | 0.012 | 0.000 | 1.000 | |
| Column Total | 341 | 53 | 31 | 425 |
| | 0.802 | 0.125 | 0.073 | |

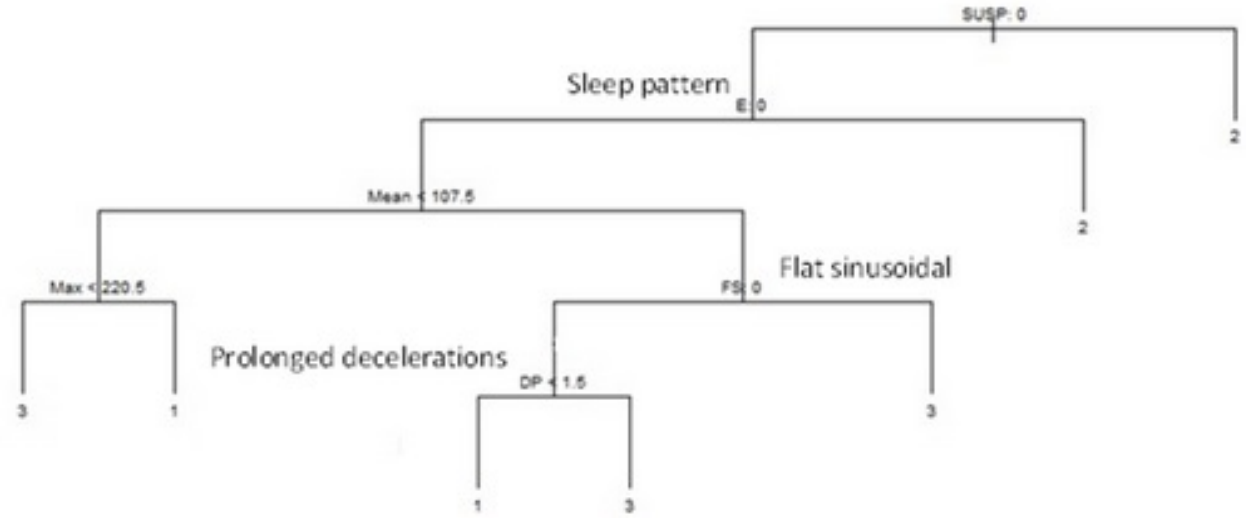



EXHIBIT P

| final_table$myresponse | final_table$predicted 1 | 2 | Row Total |
|---|---|---|---|
| 1 | 327 | 4 | 331 |
| | 0.988 | 0.012 | 0.779 |
| | 0.994 | 0.042 | |
| 2 | 2 | 92 | 94 |
| | 0.021 | 0.979 | 0.221 |
| | 0.006 | 0.958 | |
| Column Total | 329 | 96 | 425 |
| | 0.774 | 0.226 | |

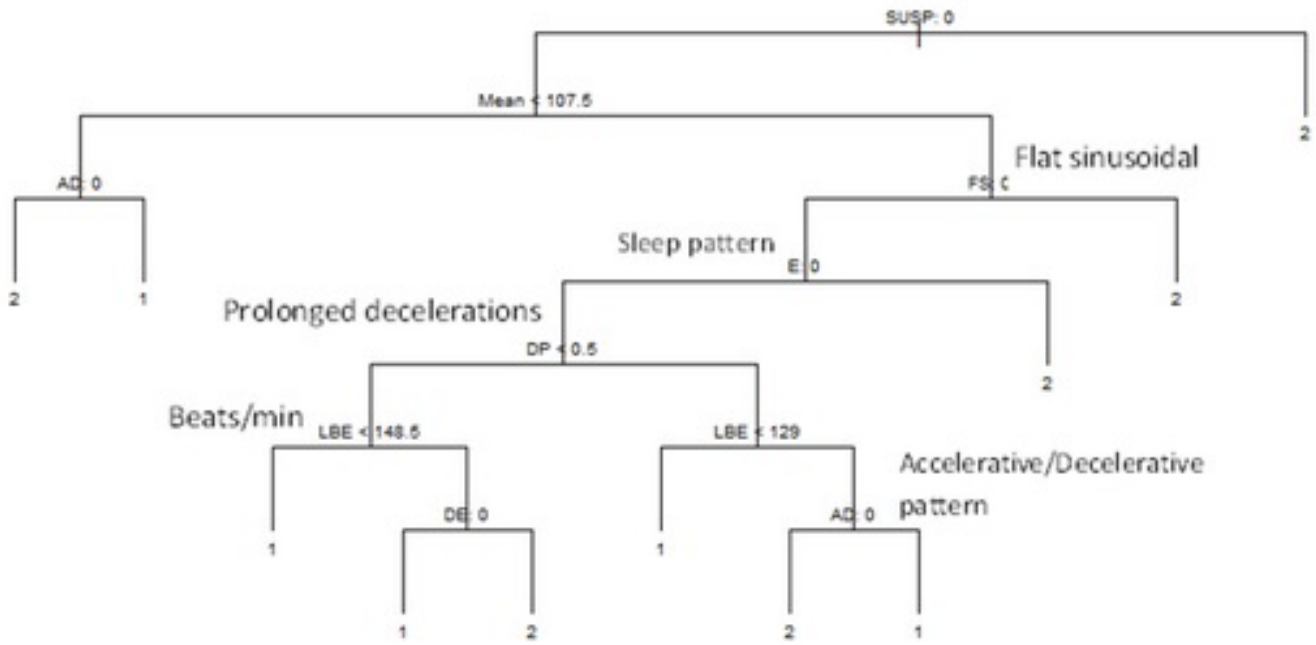



EXHIBIT Q

| True Class | 1 | 2 | 3 | Row Total |
|---|---|---|---|---|
| 1 | 330 | 1 | 0 | 331 |
| 2 | 1 | 58 | 0 | 59 |
| 3 | 1 | 0 | 35 | 35 |
| Column Total | 331 | 59 | 35 | 425 |

EXHIBIT R

| True Class | 1 | 2 | Row Total |
|---|---|---|---|
| 1 | 330 | 1 | 331 |
| 2 | 1 | 93 | 94 |
| Column Total | 331 | 94 | 425 |

EXHIBIT S

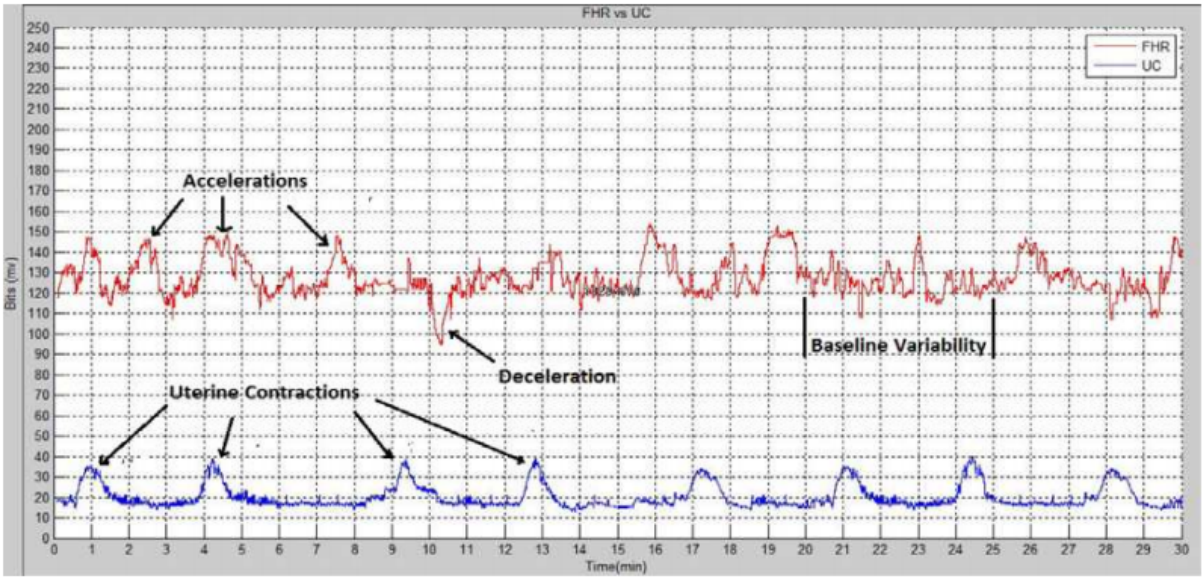



**Fig. 1.** Examples of CTG trace FHR(top) and utrine activity(bottom)

EXHIBIT T

**Multicategorical Models Classification Rates**

| Model | Normal | Suspect | Pathologic |
|---|---|---|---|
| Classification. Tree | 100% | 90% | 89% |
| KNN | 99.7% | 98.3% | 97% |

**Binary Models Classification Rates**

| Model | Normal | Abnormal |
|---|---|---|
| Classification Tree | 99% | 98% |
| KNN | 99.7% | 99% |
| Logistic Regression | 98% | 96% |

# Works Cited

Afors, Karolina, and Edwin Chandraharan. "Use of Continuous Electronic Fetal Monitoring in a Preterm Fetus: Clinical Dilemmas and Recommendations for Practice." *Journal of Pregnancy*. Hindawi, 2011. Web. http://www.hindawi.com/journals/jp/2011/848794/

Bache, K. & Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

"Cardiotocograph (CTG) Interpretation and Response". The Womens. 6. Apr. 2014. Web. https://thewomens.r.worldssl.net/images/uploads/downloadable-records/clinical-guidelines/ctg-in

"How to Read a CTG". Geeky Medics. N.p., 29 May 2011. Web. http://geekymedics.com/2011/05/29/how-to-read-a-ctg/

Marques, Juaquim P. Applied Statistics Using SPSS, STATISTICA, MATLAB and R. Berlin, Heidelberg: Springer-Verlag, 2008. Print.

Sundar, C. "An Overview of Research Challenges for Classification of Cardiotocogram Data". Journal of Computer Science 9 (2): 198-206, 2013. Web. http://thescipub.com/PDF/jcssp.2013.198.206.pdf

National Center on Birth Defects and Developmental Disabilities, Division of Birth Defects and Developmental Disabilities. July 9, 2014. Web http://www.cdc.gov/ncbddd/heartdefects/data.html